

Figure S1-A, B, C, D. Deriving search patterns for C- and H-boxes

S1A - USP C and H boxes

S1B - OTU C and H boxes

| | -8 | -7 | -6 | -5 | -4 | -3 | -2 | -1 | 0 | +1 | +2 | +3 | +4 | +5 | +6 | +7 | +8 | +9 | +10 |
|-----------|----------|----------|----------|----------|----|----------|----------|--------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| OTU C-Box | <i>H</i> | <i>C</i> | <i>S</i> | | | <i>C</i> | | | | | | | | | | | | K | |
| | <i>K</i> | <i>F</i> | <i>A</i> | | | <i>F</i> | | | | | | | | | | | R | S | |
| | <i>A</i> | <i>G</i> | <i>C</i> | | | <i>I</i> | | | | | | | | | | | C | A | |
| | <i>C</i> | <i>I</i> | <i>F</i> | | | <i>L</i> | | | | | | | | | | | F | C | |
| | <i>F</i> | <i>L</i> | <i>G</i> | <i>H</i> | | <i>M</i> | | | | | | | | | | | I | F | |
| | <i>G</i> | <i>M</i> | <i>L</i> | <i>S</i> | | <i>V</i> | | | | | | | | | | | G | C | |
| | <i>I</i> | <i>V</i> | <i>M</i> | <i>E</i> | | <i>W</i> | | | | | | | | | | | M | I | |
| | <i>M</i> | K | <i>W</i> | <i>Q</i> | | <i>V</i> | <i>E</i> | | | | | | | | | | F | C | |
| | <i>W</i> | R | <i>Y</i> | R | | <i>A</i> | <i>Q</i> | | | | | | | | | | L | T | |
| | <i>Y</i> | H | T | N | | <i>S</i> | <i>N</i> | | | | | | | | | | Y | L | |
| R | A | V | T | G | | <i>G</i> | | | | | | | | | | | G | M | W |
| | L | Y | I | D | | P | D | ^A | H | | | | | | | | L | R | V |
| | V | W | N | K | | T | A | G | D | C | L | F | M | H | A | T | S | H | V |
| | | | | | | | | | | | | | | | | | Y | T | I |

S1C - MJD C and H boxes

| | -10 | -9 | -8 | -7 | -6 | -5 | -4 | -3 | -2 | -1 | 0 | +1 | +2 | +3 | +4 | +5 | +6 | +7 | +8 | +9 | +10 |
|-----------|-----|----|----|----|----|----|----|----|----|----|---|----|----|----|----|----|----|----|----|----|-----|
| MJD C-Box | | | | | | | | | | | | | | | | | | | | | |
| A | A | A | C | G | A | C | G | F | I | M | V | A | C | F | G | I | F | A | C | A | |
| C | C | C | L | I | D | Y | H | X | C | F | G | I | M | M | L | G | I | F | C | C | |
| G | G | G | M | L | Q | H | | | G | I | I | F | I | M | G | G | F | F | G | G | |
| F | I | V | M | N | K | | | | I | M | M | L | M | I | G | I | G | G | I | | |
| M | L | W | W | H | T | A | M | | V | W | W | M | I | I | M | L | I | I | M | | |
| W | M | Y | Y | K | A | E | V | | W | Y | S | V | V | M | V | M | V | M | V | V | |
| Y | V | K | D | K | S | G | H | W | Y | N | T | W | W | W | Y | Y | Y | Y | Y | W | |
| I | W | R | Q | R | E | L | N | Y | A | L | K | Y | Y | Q | | | | | | | |
| L | F | F | N | H | N | T | R | Q | F | G | Q | R | A | L | Q | M | L | F | | N | |
| V | Y | H | E | V | Q | R | S | S | L | C | L | V | H | C | V | N | N | V | L | Q | |

| MJD H-Box | | | | | | | | | | | | | | | | | | | | | |
|-----------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|--|--|--|--|
| | | | | | | | | | | | | | | | | | | | | | |
| A | F | A | C | F | G | G | I | I | A | A | F | G | G | A | A | D | A | | | | |
| F | I | M | V | V | N | V | H | | C | C | L | C | M | F | N | C | | | | | |
| M | M | W | W | W | H | Y | K | | G | G | W | G | W | G | H | M | | | | | |
| V | V | S | Y | K | T | T | S | | M | M | Y | M | Y | M | R | W | | | | | |
| W | W | X | T | H | H | T | | | V | W | S | W | S | W | S | S | Y | | | | |
| Y | Y | F | H | H | K | K | D | | Y | Y | A | Y | Y | T | T | T | F | | | | |
| C | G | H | K | K | K | D | | | F | F | I | C | I | H | K | E | I | | | | |
| G | C | L | L | M | E | | | | I | I | T | V | V | R | Q | | | | | | |
| I | L | N | R | R | Q | | | | L | L | V | V | V | R | Q | | | | | | |
| L | P | W | Y | S | R | H | | | W | V | V | V | V | R | Q | | | | | | |

S1D - JAMM H box

Ramakrishna et al. Supplementary information

Table S1. EBV ORFs containing C- and H-box homologous regions identified by BLAST search.

| EBV ORF | C-Box ^a | H-box ^a | E-value ^b |
|--------------------|--------------------|--------------------|----------------------|
| BALF2 | OTU | | 4.5 |
| BaRF1 | | UCH | 7.1 |
| BBLF2 | | USP | 8.3 |
| BBRF3 | USP | | 0.018 |
| | UCH | | 1.4 |
| BcLF1 | | USP | 0.25 |
| | | USP | 1.6 |
| BDLF1 | | USP | 4.9 |
| BDLF2 | | USP | 8 |
| | OTU | | 3.4 |
| BDLF4 | | USP | 8 |
| BFLF2 | UCH | | 5.4 |
| BGLF4 | UCH | | 4.1 |
| BGRF1/BDRF1 | MJD | | 3.8 |
| BKRF3 | | UCH | 5.7 |
| BSLF2/BMLF1 | | OTU | 7.3 |
| BMRF1 | UCH | | 2.4 |
| BORF2 | | JAMM | 2.6 |
| | USP | | 1.7 |
| | UCH | | 1.1 |
| BPLF1 | USP | | 6.4 |
| BRLF1 | MJD | | 6.5 |
| BSLF1 | | USP | 2.1 |
| | UCH | | 9.2 |
| BXLFI | | USP | 2.2 |
| BXLF2 | | USP | 4.9 |
| ORF-13 | | USP | 8 |
| | OTU | | 10 |
| ORF-35 | | USP | 2.9 |
| ORF-38 | | USP | 6.2 |
| | | OTU | 4.3 |
| ORF-44 | | USP | 6.2 |
| ORF-88 | | OTU | 7.3 |

a. Sequence homology was identified with the conserved C- or H-boxes of the indicated DUB families. USP = Ubiquitin-Specific Protease; UCH = Ubiquitin C-terminal hydrolase; OTU = OTU domain protease; MJD = Josephin domain protease; JAMM = Jab1/MPN/Mov34 metalloprotease.

b. The E-value indicates the likelihood that the identified sequence may occur by chance in the database. An E-value of 1 corresponds to the probability that one similar sequence may occur by chance in the database.

Table S2. Putative DUB catalytic domains identified in EBV ORFs by pattern search

| EBV ORF | DUB family ^a | | EBV ORF | DUB family ^a | |
|--------------------|-------------------------|----------|---------------|-------------------------|---------------|
| | C-Box | H-Box | | C-Box | H-Box |
| BALF1 | | OTU | BNRF1 | USP | USP, OTU, MJD |
| BALF5 | | OTU, MJD | BOLF1 | | MJD |
| BcLF1 | | OTU, MJD | BORF1 | USP | |
| BDLF1 | | OTU | BPLF1 | | OTU |
| BERF1 | | OTU | BRRF2 | | OTU |
| BERF3-BERF4 | USP, OTU | | BVRF1 | | USP |
| BFRF2 | USP | USP | BXLF2 | | USP, OTU |
| BFRF3 | | OTU | BZLF1 | | USP |
| BGLF1 | USP, OTU | | LMP2 | USP | |
| BGLF4 | | OTU | ORF-06 | | OTU |
| BGRF1/BDRF1 | | OTU | ORF-35 | | USP, OTU |
| BILF1 | USP | USP | ORF-38 | | USP |
| BKRF2 | USP | | ORF-44 | | USP |
| BMRF2 | | USP, OTU | ORF-55 | | OTU |
| BNLF1 | USP | OTU | ORF-88 | | OTU |

a. Subfamily specific DUB pattern: USP = Ubiquitin-Specific Protease; OTU = OTU domain protease; MJD = Josephin domain protease

Table S3. Putative catalytic Cys and His residues in EBV ORFs conserved across HHV members

| EBV ORF | Conserved residues ^a | | HHV members | C-Box | | | H-Box | | |
|-------------|---------------------------------|-----|-----------------------|----------------------|------|-------|----------------------|------|-------|
| | Cys | His | | Residue ^b | Type | Score | Residue ^b | Type | Score |
| BALF2 | 4 | 2 | α,β,γ | | | | | | |
| BALF3 | 4 | 2 | α,β,γ | | | | | | |
| BALF4 | 9 | 1 | α,β,γ | C3 | OTU | -10.2 | | | |
| BALF5 | 8 | 4 | α,β,γ | C4 | USP | 0.3 | | | |
| BBLF4 | 3 | 4 | α,β,γ | | | | | | |
| BBRF1 | 1 | 1 | α,β,γ | | | | | | |
| BBRF2 | 1 | - | α,β,γ | | | | | | |
| BBRF3 | 1 | - | α,β,γ | C1 | USP | 7.8 | | | |
| BcLF1 | 5 | 6 | α,β,γ | | | | | | |
| BFLF1 | 8 | 2 | α,β,γ | | | | | | |
| BFLF2 | 3 | 1 | α,β,γ | | | | | | |
| BGLF2 | 5 | - | α,β,γ | C2 | OTU | -12.5 | | | |
| BGLF4 | 1 | 1 | α,β,γ | | | | | | |
| BGLF5 | - | 1 | α,β,γ | | | | | | |
| BGRF1/BDRF1 | 4 | 8 | α,β,γ | C4 | UCH | -16 | H3 | MJD | -8.1 |
| BKRF3 | 1 | 4 | α,β,γ | | | | H4 | UCH | -14.2 |
| BMLF1 | 1 | 1 | α,β,γ | | | | | | |
| BPLF1 | 1 | 1 | α,β,γ | C1 | USP | -4.5 | | | |
| BSLF1 | 6 | 2 | α,β,γ | C5C6 | UCH | -15.3 | | | |
| BXLF2 | 3 | - | α,β,γ | | | | | | |
| BXRF1 | 2 | 1 | α,β,γ | | | | | | |
| BaRF1 | 1 | 3 | α,γ | | | | H1H2 | UCH | -11.8 |
| BXLF1 | - | 1 | α,γ | | | | | | |
| BCRF1 | 1 | 1 | β,γ | | | | | | |
| BDLF4 | 5 | 1 | β,γ | | | | | | |
| BFRF1 | 2 | - | β,γ | | | | | | |
| BFRF2 | 8 | - | β,γ | | | | | | |
| BGLF3 | 4 | - | β,γ | | | | | | |
| BLRF1 | 2 | - | β,γ | | | | | | |
| BGLF1 | - | 1 | β,γ | | | | | | |
| BFRF3 | 4 | 2 | γ | | | | | | |
| BKRF2 | 4 | - | γ | | | | | | |
| BLLF3 | 4 | 2 | γ | | | | | | |
| BMRF1 | 1 | 2 | γ | | | | | | |
| BMRF2 | 2 | - | γ | | | | | | |
| BNRF1 | 2 | 11 | γ | C2 | USP | -5.0 | | | |
| BOLF1 | 2 | 4 | γ | | | | | | |
| BORF1 | - | 1 | γ | | | | | | |
| BRLF1 | 3 | 2 | γ | | | | | | |
| BRRF1 | 1 | 6 | γ | | | | | | |
| BSRF1 | 4 | - | γ | | | | | | |
| BTRF1 | 2 | 3 | γ | | | | | | |
| BVRF2 | 2 | 8 | γ | | | | | | |

a. Number of conserved Cys and His residues in EBV ORFs that are present across the HHV family members.

b. The conserved C/H residues were identified with the sequential number starting from the N-terminus.

Table S4. DUB scores of EBV ORFs from this study.

| EBV ORF | Search Strategy | | | | Score | EBV ORF | Search Strategy | | | | Score |
|-------------|-----------------|---------|-----|-------------------|-------|-------------|-----------------|---------|-----|-------------------|-------|
| | BLAST | Pattern | HMM | Conserved Cys/His | | | BLAST | Pattern | HMM | Conserved Cys/His | |
| BBLF3 | 3 | | 3 | 3 | 9 | ORF-38 | 2 | 1 | | | 3 |
| BALF5 | | 1 | 3 | 3 | 7 | BBLF4 | | | 2 | | 2 |
| BcLF1 | 3 | 1 | 3 | | 7 | BBLF2 | | | 2 | | 2 |
| BGRF1/BDRF1 | 2 | 1 | 2 | 2 | 7 | BGLF3 | | | 2 | | 2 |
| BNRF1 | | 2 | 2 | 2 | 6 | BKRF1 | | | 2 | | 2 |
| BPLF1 | 1 | 1 | 2 | 2 | 6 | BKRF3 | 1 | | | 1 | 2 |
| BALF2 | 2 | | 3 | | 5 | BMRF1 | 2 | | | | 2 |
| BSLF1 | 2 | | 2 | 1 | 5 | BMRF2 | | 1 | 1 | | 2 |
| BXLF2 | 2 | 1 | 2 | | 5 | BSLF2/BMLF1 | 1 | | 1 | | 2 |
| ORF-35 | 2 | 1 | 2 | | 5 | BVRF2 | | | 2 | | 2 |
| ORF-88 | 1 | 1 | 3 | | 5 | ORF-101 | | | 2 | | 2 |
| BaRF1 | 1 | | 2 | 1 | 4 | ORF-13 | 2 | | | | 2 |
| BREF3-BERF4 | | 1 | 3 | | 4 | ORF-44 | 1 | 1 | | | 2 |
| BILF1 | | 2 | 2 | | 4 | ORF-86 | | | 2 | | 2 |
| BRLF1 | 1 | | 3 | | 4 | BcRF1 | | | 1 | | 1 |
| BXLF1 | 2 | | 2 | | 4 | BDLF4 | 1 | | | | 1 |
| BALF1 | | 1 | 2 | | 3 | BERF1 | | 1 | | | 1 |
| BALF4 | | | 2 | 1 | 3 | BFLF2 | 1 | | | | 1 |
| BBLF2 | 1 | | 2 | | 3 | BFRF3 | | 1 | | | 1 |
| BDLF1 | 2 | 1 | | | 3 | BKRF2 | | 1 | | | 1 |
| BDLF2 | 2 | | 1 | | 3 | BLLF3 | | | 1 | | 1 |
| BFRF2 | | 2 | 1 | | 3 | BOLF1 | | 1 | | | 1 |
| BGLF1 | | 1 | 2 | | 3 | BRRF2 | | 1 | | | 1 |
| BGLF2 | | | 2 | 1 | 3 | BVRF1 | | 1 | | | 1 |
| BGLE4 | 2 | 1 | | | 3 | BXRFL | | | 1 | | 1 |
| BNLF1 | | 2 | 1 | | 3 | BZLF1 | | 1 | | | 1 |
| BORF1 | | 1 | 2 | | 3 | ORF-06 | | 1 | | | 1 |
| BORF2 | 3 | | | | 3 | ORF-42 | | | 1 | | 1 |
| LMP2 | | 1 | 2 | | 3 | ORF-55 | | 1 | | | 1 |

a. Scoring criteria:

BLAST: **0** = domains with E-value >10 or lacks C- or H-box; **1** = C- or H-box homology domains with E-values ≤10 and ≥5; **2** = C- and H-box homology domains with E-value ≥5 or one C- or H-box homology domain with E-value between 5 and 1; **3** = C- and H-box homology domains with E-value value between 5 and 1 or one C- or H-box homology domain with E-value ≤1 **4**= C- and H-box homology domains with E-value value ≤1

Pattern search: **0** = no domain with C- or H-box homology found; **1** = presence of domain(s) with either C- or H-box homology; **2** = presence of both C- and H box homologous domains

HMM search: **0** = no domain with C- or H-box homology found; **1** = either C- or H-box homologous domain with HMM score ≤-6.6; **2** = both C- and H-box homologous domains with HMM score ≤-6.6 or presence of either C- or H-box homologous domains with HMM score between 0 and -6.6; **3** = presence of both C- and H-box homologous domains with scores between 0 and -6.6 or presence of one domain with HMM score ≥0; **4** = presence of both C- and H-box homologous domains with HMM scores ≥0

Conserved Cys and His search: **0** = no conserved Cys/His residue or conserved residues in regions with no homology with C- or H-boxes; **1** = one conserved Cys or His residue in a C- or H-box homologous domain with score <-10; **2** = conserved Cys and His residues in C-and H boxes with scores <-10 or one conserved Cys or His residue in a C- or H-box homologous domain with score between -10 and 0; **3** = conserved Cys and His residues in C- and H-box homologous domains with scores between -10 and 0 or one conserved Cys or His residue in a C- or H-box homologous domain with score >0; **4** = conserved Cys and His residues in C- and H-box homologous domains with scores >0.